

Rosa Rabadán

Belén Labrador

Noelia Ramón

University of León, Spain.

ENGLISH-SPANISH CORPUS-BASED CONTRASTIVE ANALYSIS: Translational Applications from the ACTRES¹ Project.

1. The framework: DTS, corpus-based research and “applied extensions”

In the last couple of decades there has been an unprecedented development in Translation Studies in line with findings in related fields. This has meant that for many, translational research has moved away from language and has concentrated heavily on its dimension as a culture transfer phenomenon. The truth, however, remains that translation and translating are basically and primarily both a textual linguistic operation and the result of such a process. Exceptionally powerful models, such as Descriptive Translation Studies (DTS), do incorporate this dimension in their epistemological framework, but for some reason or other, most research has focused on the study of translational behaviour regularities, the reception of the translated product and the target-context consequences of the types of transfer observed, leaving the painfully compiled corpora unexploited further than their value as a catalogue for deriving preliminary norms.

Other paradigms, such as corpus-based research, which is intimately connected with DTS, are specifically geared to the search of the so-called universals of translation, that is, to define the specific features of translation-produced language such as simplification, explicitation, conservatism, etc. (Laviosa 2002: 43-55). This means that there is a gap, an applied gap, in the enquiry, and that it has had further implications, namely the existence of a breach between those who see translation just as action and therefore need prospective information about operational norms, and those that have not been able (or have been unwilling) to derive this prospective information from retrospective descriptive research, and this includes data collected by means of translation memories. From this point of view, contrastive analysis follows a parallel strategy to that of DTS, and the formulation of regularities springing from a bilingual comparable corpus would amount to an **applied extension** in Toury's terminology (1995). While sharing procedures and research protocols that are properly described as descriptive, contrastive analysis emerges as the most powerful basis for translation applications as it provides the necessary data for translator aids, translation training and assessment. Attempts at gaining significant aid from corpora when translating largely remain wishful thinking (owing mostly to the time-consuming nature of searches) and for the most part reduced to knowledge retrieved from the user-friendly translation memories, the use of corpora in translation training is still in need of further development, as shown by the works of Ulrych (1997), Bowker (1998) and Zanettin (1998), and translation quality assessment has long been identified as one missing component in the DTS framework, as noted by advocates of functional models (Nord 1997). This, however, does not mean that corpora do not have much to contribute to the development of TS, especially in the line of integrating the findings of linguistic and cultural approaches to translation. What actually happens is that language-pair specific applications have not yet been developed.

2. The interface between DTS & CA: how to address it

Our own work in TRACE², a long-term corpus-based project that studies censored translations in the Franco era using parallel corpora, showed that descriptive work at the textual and linguistic lower

¹ *Contrastive Analysis and Translation English-Spanish*, funded by MCyT BFF2001-0112 and JcyL CO03/102.

² For state-of-the-art findings concerning this project see Rabadán (2000).

levels, was far from being properly researched, let alone the ways and tools to proceed with language choice assessment. Quite naturally, the answer came in the form of applied corpus-based contrastive analysis. Retrospective research along DTS lines aims at unveiling the regularities in translation behaviour across a corpus, which obviously includes regularities in language transfer, recurrence of solutions for the same translational problems, and in most cases it aims at interpreting these regularities as norms bearing witness to translation as a factor of cultural exchange and cultural dynamics. However, it is difficult to decide upon a possible interpretation of translation produced language if the typical non-translated native usage is not brought into the picture.

On the basis of this very basic premise, that descriptive results -of whichever kind- at the language level obtaining from parallel corpus research should be confronted with actual native usage before attempting any serious interpretation, the ACTRES project set out some years ago with two clear aims, i) identifying the expressive resources available in both English and Spanish for one given semantic function by using comparable corpora and ii) establishing the translational options under different textual and pragmatic circumstances. These options, chosen according to their regularity of occurrence, i.e. their typicality, would be formulated as norms 'with an option'. Their applicability range is wide and ranges from the primary premise of providing a descriptive *tertium comparationis* for the interpretation of translational shifts obtained from work done with parallel corpora, to the formulation of instruction-like multiple choice translational options to be used by practitioners and translation trainers.

3. The ACTRES project: stages and procedures

ACTRES is organized on levels of analysis and methodological phases. Work has been distributed and taken up by team members according to their specialization in one of three levels: grammar, rhetoric and discourse analysis and lexicology and terminology.

Contrastive rhetoric aims at establishing the differences between discourse structures across cultures, languages and genres (Connor 1996:14&ff). In ACTRES, and because of the project's outlook on translational applications, this includes the pragmatic application of each genre in the corresponding context of situation. What is contrasted in ACTRES are texts that have the same intended communicative goals in their respective cultural and linguistic contexts (Trosborg 2000). The aim of research can be arrived at from three different angles, by contrasting: 1) the rhetorical structure of texts sharing the same communicative function in English and in Spanish; 2) macrostructures that contribute the same functional meaning (i.e. do the same job) within each text, or 3) the actualisation in each language and text of the functional-semantic dynamic parameters (i.e. intended recipients, text mode peculiarities, intertextual adjustments, etc). It has been proved that combining the three foci in the same study produces substantial increase in the significance and applicability of the findings. Rhetorical research from the ACTRES project makes use of ready-made corpora in order to explore the textual output of different areas of knowledge. So far, two main areas and their sub-areas have received our attention, business and finance and the biosciences, specifically biomedicine and biotechnology.

Contrast at the lexicographic levels in ACTRES concentrates on bilingual terminology of specialized fields (biomedicine, finance, wine production, etc). The tools used are ready-made bilingual comparable corpora built to the specifications of a term bank such as TERMIUM³, which will eventually incorporate the results into its files. Because of its peculiarities, it is an area with clearly defined aims and research protocols within the ACTRES project, and although it shares the same basic theoretical and methodological assumptions as the other two levels, the exploitation of its findings can be undertaken independently.

Grammatical contrast concentrates on those areas of meaning or on those formal resources that are a source of recurrent problems when translating from English into Spanish. The underlying language conception is Bondarko's functional-semantic fields (1991) where one core meaning is represented primarily by one formal resource, which is the typical central one, and by a varying number of other expressive means which are more or less peripheral according to a typicality scale, a conception which can be seen as parallel to that of translational regularities in DTS.

³ <http://www.termium.com>

The standard procedure follows four steps: 1. identification of a problem area and reasons for it being so (our selection phase); 2. description of meaning function and typical expressive resources both in English and in Spanish; 3. juxtaposition of expressive means and typicality scales in both languages; and 4. contrast and possible correspondences.

Our first phase is initiated using one of two procedures: either we use knowledge gained from translational practice (e.g. adjectival strings; modals; imprecise quantification, etc.) or we may require more specific information and use evidence from a parallel corpus of original English texts and their translations into Spanish as a diagnostic tool. When dealing with meaning areas the analysis goes from meaning to form in each of the languages concerned and from local forms to shared meanings, so as to establish the map of potential translational solutions. Whereas if we are concerned with polyfunctional resources in the SL (e.g. -ing English forms), the analysis will proceed from form to meaning in each of the languages via a parallel corpus, and then on from shared meanings to acceptable and correct native usage. When we adopt this second procedure, the translated solutions offered by the parallel corpus are taken as an input to the Spanish part of the comparable corpus in order to check whether they are taken to mean the same in Spanish or are used differently.

The second stage, the description, based on previous grammatical work done along functional lines, will make it feasible to produce an outline of functional meanings and correlated forms in both languages. Functional classifications, when they exist, have proven extremely useful, as they provide a descriptive grid for systematically checking interlinguistic divergence. By the end of this stage we can count on a map of expressive possibilities in both languages and their contextual applications.

The third step, juxtaposition, is parallel to comparative analysis in DTS, and it consists in systematically observing regularities in use in both languages concerning the expression of one given meaning/function. Quantitative data are qualitatively interpreted and hypothesis concerning possible usage correspondence between both languages are put forward. Combinatorial restrictions affecting meaning are also taken into account and integrated in the hypotheses.

And finally, the stage we have dubbed contrast, which runs parallel to the formulation of translational norms in DTS. The contrast phase is the input of the actual **application to translation**: a collection of available translational solutions to one given translation problem, choice restrictions included. At this point instruction-like regularities can be formulated in the form of (restricted) multiple choice- e. g. for translational problem A, you have the following possible solutions: X (best when...), Y (preferable if...), Z (only to be used if...).

4. Corpora applications English-Spanish: COBUILD-Bank of English and CREA

In order to establish these potential translational options we need real life observational data from which to derive our proposals. Our primary tools in ACTRES are two large monolingual corpora, one English, the other Spanish, featuring an equivalent internal architecture concerning subcorpora, intralinguistic varieties and statistical dimensions. Corpus selection started from the very basic condition of availability in both languages. For Spanish, there is certainly a range of possibilities, although the one best suited to our purposes is the CREA⁴, sponsored by the Spanish Royal Academy for Language; and this choice determined the selection of the English corpus, which for reasons of equivalence is Cobuild's Bank of English.⁵

As of today, the Bank of English consists of approximately 415 million words, 56 of which are available online through *CobuildDirect*. It comprises 12 different full-text subcorpora, selected according to geographical provenance, textual mode, and physical format. Practically all materials date from 1990 onward with constant updating and they are tagged for part-of-speech. Search tools are integrated in the corpus, which makes it user-friendly.

CREA consists of approximately 130 million words and is also a full-text corpus. The criteria for text selection are comparable to those of Cobuild: geographical provenance; textual mode; physical format, and an extra bonus: field area, which can prove immensely helpful in certain types of research (e.g. phraseology). CREA is also subject to continuous updating, keeping only those texts produced in the last 25 years, which means that regularly part of the older materials are moved to the diachronic CORDE database. Searching CREA means using a lexical input, as part-of-speech tagging is not yet complete. This difference in software tools, together with typological and distributional differences have forced us

⁴ Corpus de Referencia del Español Actual (CREA): <http://www.rae.es>

⁵ Online access via : <http://titania.cobuild.collins.co.uk>

to come up with a variety of searching strategies, all geared towards one unique aim - extracting ways of naturally expressing one meaning in English and in Spanish.

Concerning corpus comparability, the subcorpora chosen are, for both corpora, those comprising written texts: newspapers, magazines, books and ephemera. The chronological span in both Cobuild and CREA is from 1990 to 2001, with figures of approximately 30 m words for the English corpus and 36 m words for the Spanish one, and the language variety choice is in both cases the European one. The only single criterion which is available and has not been used as a filter is what we have called “field/topic area”, as it is only applicable in CREA.

5. Case studies

The procedures outlined in 3 above were applied to two specific semantic areas which are frequently branded as a source of translational problems – the characterisation of nouns and the expression of indefinite quantity. Throughout both studies form and meaning are taken as the input criteria at different stages. The selection phase is onomasiological, as it goes from meaning to form – from characterisation and quantification to the lexico-grammatical resources employed by native speakers of English and Spanish to actualise those semantic functions. The description stage is semasiological, as it starts from form in order to assign particular semantic functions to each occurrence. Juxtaposition and contrast are onomasiological also, as the aim is to establish the various possible formal realisations in each language for the same meanings.

5.1. Selection

5.1.1. Data selection in the study on quantification

For the study on quantification, one particular class, indefinite quantification, was chosen as the object of the empirical contrastive study. First, a list of English and Spanish indefinite quantifiers was elaborated using a number of English and Spanish reference grammars. The author’s own intuition and the opinion of several native informants also played a role in selecting the formal input. The indefinite quantifiers studied amounted to 188. Their frequency rates in CREA and Cobuild’s Bank of English were then searched and a statistical formula was applied in order to obtain a representative number out of the whole population of concordances for each quantifier.

5.1.2. Data selection in the study on characterisation

The need for formal input in a corpus-based analysis led us to use nouns as semantic nodes in the case of nominal characterisation, so that the syntactic surrounding of these nouns could be analysed. Frequency lists of nouns were checked in both languages and the ten most common nouns in each case were selected to carry out the analysis. A statistically representative number of concordances of these nouns was downloaded from the two corpora, and the head nouns were searched without any further specification in Spanish, and using part-of-speech tags in English so as to avoid the occurrence of verbs that have the same form as nouns.

5.2. Description

5.2.1. Descriptive stage in quantification

Once the data were downloaded, each of the concordances was classified according to semantic criteria and manually tagged for function and pattern. The quantifiers were divided into seven groups parallel in the two languages according to the following characteristics:

- Negative quantifiers generally express lack of quantity, e.g. *none, ninguno*.
- Universal quantifiers indicate the totality of a quantity in a concrete universe, restricted on many occasions by adjectives, relative clauses, non-finite clauses, etc., e.g. *all, todos*.
- Existential quantifiers refer to the existence of some quantity as opposed to absence and totality, e.g. *several, algún*.
- Proportional quantifiers are divided into multal, paucal and relational. Proportional multal quantifiers express a high amount of something, higher than the assumed norm along a scale, e.g. *much, mucho*; proportional paucal quantifiers express a low amount, lower than the assumed norm, e.g. *little, poco*; and proportional relational quantifiers (e.g. *enough, bastante*) express an amount that is directly related to the upper and lower parts of the scale and also very closely

linked to two other functions: excess and sufficiency. Thus, relational quantifiers do not express amounts that are large or small according to a proportional scale but on the basis of a particular need.

- Relative quantifiers express a type of quantity either unknown, as is the case of interrogative pronouns, indefinite, as in the forms ending in *-ever* (*whatever*), or *-quier/a* (*cualquier*), or non-assertive as in *any* and its compounds.

5.2.2. Descriptive stage in characterisation

After the process of data selection, the instances in which the head noun was characterised by one single modifier were identified and classified. In addition to adjectives, modifiers included nouns, adverbs, adjective phrases, prepositional phrases, relative clauses, -ed clauses, -ing clauses, appositions, etc.

Semantic classifications of adjectives were used as a starting point and were then equally applied to all modifiers. The most widely used semantic typology distinguishes between descriptive and classifying meanings (Ferris 1993; Biber et al. 1999). Adjectives with a descriptive meaning add a quality to the head noun that is not an intrinsic part of the nature of that noun, as in *a dangerous man*, *un problema grave*. These adjectives may appear in predicative position (*the man is dangerous*) and are susceptible of being modified (*un problema muy grave*). In contrast, adjectives with a classifying meaning “delimit or restrict a noun’s referent, by placing it in a category in relation to other referents.” (Biber et al. 1999: 508), as in *a polar bear*, *una situación social*. Classifying adjectives occur exclusively in attributive positions (before the head noun in English, and after the head noun in Spanish), with no other modifier interfering, and categorise the head noun.

However, the inclusion of a wide range of grammatical resources required a wider semantic classification, and semantic headings such as time, place or manner, were added to the basic distinction between descriptors and classifiers.

5.3. Juxtaposition

This third step in the two case studies consisted in matching each of the functions identified with the range of formal patterns used to express them, both in English and in Spanish. The data were juxtaposed in the form of tables with two columns: the English resources found for expressing a particular meaning were listed on the left-hand side, and on the right-hand side the Spanish resources found for expressing the same meaning. In both columns the resources were ordered according to their frequency of occurrence, thus distinguishing the most typical constructions from those more peripheral.

5.4. Contrast

The final contrast involves putting forward functional-semantic equivalents that can be considered translational choices for one particular meaning.

5.4.1. Contrastive stage in quantification

Altogether 56 functions expressed by quantifiers were identified. A distinction was made between those that are closely related to quantification, 33, e.g. ‘intensification’ – see table below for the central patterns - and those bearing scarce relation to quantification, 23, e.g. ‘concession’, as in *after all*. The quantifying functions account for approximately three quarters of all the occurrences.

English		Spanish
premodification with multal and paucal quantifiers (e.g. ‘an awful lot’, ‘so many’, ‘a tiny little bit’)	Intensification	suffixes (e.g. ‘-ísimo/a/os/as’)
repetition (e.g. ‘loads and loads’, ‘masses and masses’)		relative quantifiers (e.g. ‘tanto/a/os/as’, ‘cuánto/a/os/as’)

Table 1. Prototypical uses of English and Spanish to express intensifying quantification.

5.4.2. Contrastive stage in characterization

An additional way of presenting contrastive results was devised in the form of inverted pyramids, such as the following one, which represents the data corresponding to the classifying function in English and Spanish.

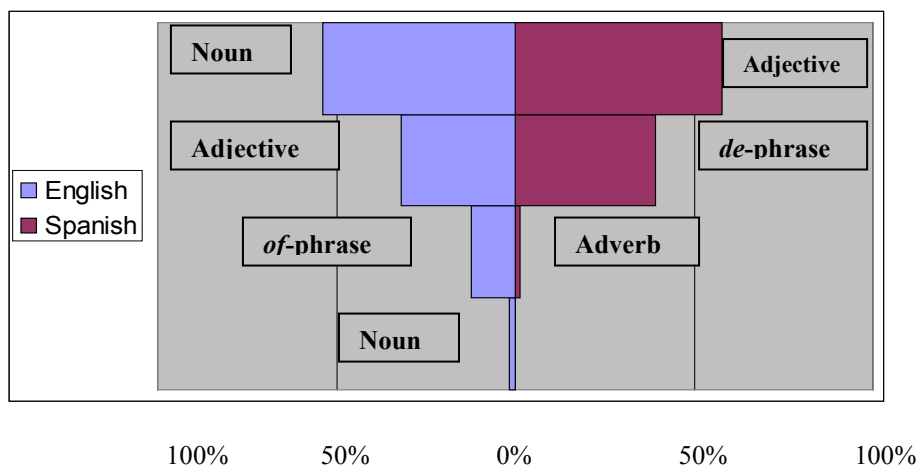


Figure 1. English-Spanish resources for the classifying function.

Our data have revealed that the two most central and typical classifying resources in English are nouns and adjectives, in that order. Other resources may be used (*of*-phrases and adverbs), but they are only marginal. The functional-semantic equivalents in Spanish will be the most typical modifiers found in that language, namely adjectives and *de*-phrases.

6. Conclusion

Corpora applications along contrastive lines have proven to significantly improve translational performance, as they provide a reliable tool for systematization as well as self-assessment. Further “applied extension” analyses will hopefully yield the map of English-Spanish translational options for those problem areas.

References

- Biber et al. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Bondarko, A. V. (1991). *Functional Grammar. A Field Approach*. Amsterdam/ Philadelphia: John Benjamins. [Russian original, 1984]
- Bowker, L. (1998). “Using specialized monolingual native-language corpora as a translation resource: A pilot study”. In S. Laviosa (ed). 631-651.
- Connor, U. (1996). *Contrastive Rhetoric: Cross-cultural Aspects of Second Language Writing*. New York: C. U.P.
- Ferris, C. (1993). *The Meaning of Syntax. A Study in the Adjectives of English*. London: Longman.
- Laviosa, S. (2002). *Corpus-based Translation Studies*. Amsterdam y Nueva York: Rodopi.
- Nord, C. (1997). *Functional Models of Translation*. Manchester, UK: St Jerome.
- Rabadán, R. (ed). (2000). *Traducción y censura inglés-español 1939-1985. Estudio preliminar*. León: Universidad de León.

Toury, G. (1995). *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins.

Trosborg, A. (ed). (2000). *Analysing Professional Genres*. Amsterdam/Philadelphia: John Benjamins

Ulrych, M. (1997). "The impact of multilingual parallel concordancing on translation". In *PALC' 97. Practical Applications of Language Corpora*. Łódź: Łódź University Press. 421-435.

Zanettin, F. (1998). "Bilingual comparable corpora and the training of translators". In S. Laviosa (ed). 616-630.

KEYWORDS: ACTRES, contrastive analysis, applied extension, translation studies, bilingual comparable corpora.